PLANNING FOR AN INSTITUTIONAL REPOSITORY: LESSONS LEARNED AT OREGON STATE UNIVERSITY

Janet Webster
Guin Library, Hatfield Marine Science Center
Oregon State University
Newport, OR 97365 U.S.A.
E-mail: janet.webster@oregonstate.edu

ABSTRACT: The Oregon State University Libraries began working on implementation of an institutional repository through a five-month planning process. The Task Force responsible for the planning identified campus needs for such a service, examined persistent issues and challenges, and made recommendations for a pilot project. These steps are summarized and lessons learned articulated.

KEYWORDS: Institutional repositories; digital libraries; scholarly communication; metadata

Introduction and Background

Institutional repositories create a virtual and intellectual environment for the digital output of a defined community such as a university or a discipline. They are an attempt to address the challenges of digital archiving, the expectations of the campus community for better access to information, and the inadequacies of the current cumbersome model for scholarly communication. There are various organizational models, software and hardware being developed and implemented. In general, the approach is still new and evolving. The OSU Community develops electronic resources constantly, and needs a means to archive and distribute many of them appropriately. One option is to implement an institutional repository (IR). In September 2003, Karyle Butcher, University Librarian at Oregon State University (OSU), appointed a task force to explore institutional repositories as one mechanism to address the challenge of collecting, maintaining and serving the digital output of OSU.

Careful planning is warranted as implementing an IR represents a significant investment in time and effort as well as being a new approach to collections and services. Consequently, the Task Force spent several months interviewing colleagues at other institutions, surveying peer institutions, reading extensively, and talking to prospective partners and users of an IR on campus. We identified specific as well as broad needs at OSU for capturing, storing and providing access to digital content. Peers helped us articulate persistent issues with implementing and maintaining IRs. Initial and ongoing costs were projected. The final report of the Task Force covers these points and recommends a pilot project using DSpace (Webster *et al.* 2004).

The following summarizes the selected needs of this particular academic community and describes the persistent issues identified through the planning process. The report's recommendations are also included to give others considering such a venture one model for planning.

Identifying the Needs of the OSU Community

This is a moving target as content is created daily in new forms and by varying members of the campus community. The Task Force spent time identifying who was currently creating digital content and what content was consistently or historically difficult for the library to collect. From this, we developed a list of people, departments and campus services units to interview. The needs described below are only a sample of those we identified.

- Need for improved archiving of department publications
  Departmental working series and technical papers have always been a challenge to collect consistently. As many have transitioned to digital format, the established acquisition workflow is disrupted and the university loses track of important institutional documents. An IR would potentially broaden access, reduce staff time and free space from redundant storage.

- Need for improved access to faculty research papers
  This need is both practical and philosophical. The practical aspect is that while many faculty members self-archive via personal web sites, there is not a coherent means of searching or managing these distributed archives. The philosophical aspect is perhaps more abstract, but addresses the greater good of improving the scholarly communication landscape through open access. According to Peter Suber, "The public interest lies in open access (OA) because open access shares knowledge, accelerates research, and multiplies all the benefits of research" (2004). Clifford Lynch describes the IR "as a new strategy that allows universities to apply serious, systematic leverage to accelerate changes taking place in scholarship and scholarly communication" (2003). It behooves the institution to promote its research output in a more accessible manner.

- Need to archive faculty datasets and databases
  We are concerned with losing valuable syntheses of information in faculty-developed databases and websites as people retire or move onto new projects. In the past, some of these would have been published as monographs; now, faculty members look to the library for guidance in preserving the content. These are problematic items to collect as the software is often non-standard, and the interfaces varied.

- Need for improved access to progress and final research grant reporting
  One way to promote the university is to showcase its research output in a coherent manner. Of course, published papers and monographs provide glimpses of this. However, compiling the research grant award and outcome information can create a compelling snapshot of the university's output and impact on society

- Need to capture and store theses and dissertations electronically
  People both internal and external to the university regularly request digital access to OSU theses and dissertations. Other institutions use the IR as a means of facilitating the theses process from review through archiving of the final copy.

- Need to capture undergraduate research accomplishments
  Increased emphasis on the student experience at OSU and more opportunities for undergraduate research leads to more interest in tracking what these students are doing. Examples of student work provide a perspective on what students learn and how they communicate that learning

- Need for better control and access to born-digital photographic images
  Since 2001, OSU communications offices have relied on digital photography to produce images for a variety of purposes including illustrating news releases and both internal and external publications. The resulting accumulation of images creates challenges in the storage, access and preservation of these potentially historically significant resources

Identifying Persistent Issues

Key points, issues and challenges emerged from conversations with peers in the United States and well as from the growing body of published reports and articles on IRs. Given the amount of discussion, relatively few institutions have implemented an IR. Some European institutions have longer experience with the service than those in the United States. Yet, the issues are consistent for all planning an IR.

- Participation by faculty/researchers and the instituion
  Colleagues all mention the importance of faculty or researcher participation and in the next breath, the difficulty of garnering that participation. Achieving critical mass, while important, is proving difficult at almost all IRs (SPARC & Crow 2002). It is necessary to understand the needs of various audiences and then keep promises made concerning access and permanence.

- Rights management
  Faculty members have concerns about ownership of repository items, pre-publication, and withdrawal rights (SPARC & Crow 2002). Tension can develop between a library's desire to retain all versions of items for historical record versus the faculty members' wish to present only the most recent or accurate version. There

is some concern about copyright infringements when control of submission is left to faculty groups and departments.

- Quality assurance of metadata
  Generating useful metadata while encouraging distributed effort poses a significant challenge to all interviewed. To some libraries, there is a control issue, a reluctance to give over collection and organization decisions. However, most are concerned with how to maintain the quality of the metadata so findability is preserved and the library does not have to mediate every entry. "Standardized metadata is central to interoperability; at its best it is a powerful tool that enables the user to discover and select relevant materials quickly and easily. At worst, poor quality metadata can mean that a resource is essentially invisible within a repository or archive and remains unused" (Barton *et al.* 2003). Additional work is required to define content rules, improve metadata entry tools, and implement a quality control process. This is doable, but not simple.

- Workflow & Integration with Existing Digital Projects
  While metadata is critical, it is only one part of the workflow required for a vital IR. Those interviewed tend to focus on the role of faculty as contributors without adequately defining the role of the library. It seems critical to describe the vision and scope of the IR as well as the role of the library versus other partners. This will help shape the workflow as well as potentially integrate the IR with existing digital projects from a functional perspective. William Nixon, University of Glasgow, describes five roles in addition to administration and systems management that the library should assume for a successful IR (2002):

    - Encouraging members of the University to deposit material;
    - Providing advice to members of the University about copyright and rights issues;
    - Converting material to a suitable format;
    - Depositing material directly on behalf of members of the University who cannot;
    - Reviewing the metadata of content.

  All require discussion about level of library involvement as well as assigning responsibility. Workflow and integration issues, while not trivial, are manageable if addressed early in planning, if roles are assigned thoughtfully, if resources are made available and if expectations are explained.

- Archival/Preservation commitment
  All would agree that a major selling point of an IR is the sense of permanence. However, all would also agree that the technology is not there to guarantee permanent archiving of all digital formats. Cornell, a leader in digital archiving, has created a 'trusted digital repository' model that integrates and highlights the importance of preservation in repository development (McGovern 2003). At this

point in time, the most constructive approach is to be aware of the issue, be involved in discussions of solutions, and adhere to standards.

- Persistent Identifier commitment
  The persistent identifier is another piece of the commitment to permanence. DSpace uses the CNRI handle system, a widely recognized standard. The Digital Object Identifier (DOI), increasingly common in digital publishing, uses the Handle system. Basically, this system assigns, manages and resolves persistent identifiers. A digital item is assigned a handle (e.g., a Uniform Resource Indentifier or URI) that relates to current information about the item and its location. The information can change, but the handle remains the same, so the item has a persistent location (CNRI 2003). This is a critical component of an IR and one that all products appear to incorporate.

- Migrating, or "the exit strategy"
  Rarely explicitly mentioned, this is something to consider when implementing an IR and especially if contemplating extensive customization of the records or metadata. Staying content-centric and applying standards and protocols that are widely used are encouraged (SPARC & Crow 2003).

- Policy Development
  Once the novelty of an IR's possibilities fades, the policy issues persist. General policy decisions are made early such as scope and contributors. Scope refers to the overall content. It is imperative to set broad guidelines initially so participants have a grasp of vision and scope. It is also important for individual communities of users to have flexibility in establishing their own content and use guidelines. Generally, policies fall into three categories: submissions, removals, and rights and permissions. Coherent policies will affect the usability of IRs as much as open source software and quality metadata. Clifford Lynch makes an excellent case for avoiding policies that become control devices and blurring responsibilities for scholarly communication. He admonishes us to "respect institutional repositories as infrastructure and not overload this infrastructure with distracting and irrelevant policy baggage" (Lynch 2003).

- Technical support
  Concerns here are threefold: installation, ongoing maintenance/operations, and support for submitters. None is insurmountable, but how they are addressed will affect the cost and success of the IR.

- Storage
  Storage requirements will depend on type of material (e.g., PDFs versus streaming video) and number of items. Electronic theses and dissertation storage requirements suggest an average of 1mg per thesis with 5-10 for those with images or data models (Fox *et al.* 1996). Images would have greater storage demands. In the long-term, there is an issue of who is responsible for the supplying the storage if it is truly a shared institutional resource

- Costs
  While most IR software is free, an IR is not. For example, implementation of DSpace is estimated to cost between $10,000 and $50,000 depending on equipment choices and level of customization (PALS 2004; Barton & Walker 2002). Ongoing costs include staff, systems support, and equipment reserves. Staffing costs break into two major categories: project management and system support. Costs are proportional to the performance and fault tolerance levels the institution desires (SPARC & Crow 2002). Costs will also vary depending on the amount and type of material deposited; more storage and higher performance needs will drive costs higher. Project management costs include project coordination, training users, marketing, metadata creation and quality control, and possible digitization or format conversion. All costs are dependant on the level of activity and the commitment of the institution. If possible, the library should secure institutional support through funding and its tacit recognition of value. Without new funds, the IR will be limited in scope and utility.

Articulating Recommendations:

Once needs were identified and issues explored, the Task Force recommended a pilot implementation of DSpace to be followed by a period of experimentation and seeding of the repository in partnership with identified units on campus. We choose DSpace as our IR software based on its growing user base, adequate functionality, and community-based structure (DSpace Federation 2004). It appears to be a manageable technology given our capacity and level of expertise.

We developed a vision for the IR so all involved had a simple statement about what the IR could be.

> "As one tool in the OSU Libraries' suite of digital library tools, the OSU's Institutional Repository will provide a reliable means for faculty members to store and access their research and teaching output, for students to do the same with their research, and for the institution to maintain its historical record" (Webster *et al.* 2004).

We also outlined roles for different parts of the Library during the pilot project. Library Technology would gain familiarity with the hardware, software and performance needs of the system. The Library Faculty would learn how to set up and run a community within the IR. Technical Services would develop a workflow to assist in the depositing of materials. All the above would assist in efforts to market the IR campus-wide.

The Task Force drew up initial steps with a timeline and assignment of responsibility. For example, the installation of the software was given to the Library Technology unit with a suggested time frame for completion. We also communicated additional specific tasks that we thought needed to be addressed for the success of the pilot project. Finally, we

recommended a review of progress with the intention of soliciting university funding for expansion of the initiative.

Lessons Learned:

Planning can be tedious and time-consuming, especially when you have the impression that everyone around you has figured out the issue and moved towards implementation. The buzz around IRs has been increasing over the past several years and library administrators can be tempted to start one just to stay current. Yet, patience can be useful. Early adaptors are sharing their challenges and issues with implementing and maintaining IR. Listening carefully is as important as being on the cutting edge.

The OSU Libraries' planning process appears to have been useful. We listened to the wisdom and advice of colleagues so we are better prepared for the pitfalls and challenges. We have initiated conversations with our local community of users. From those conversations, needs emerged that we may be able to address with an IR. People outside the library are now getting interested in the project. The groundwork is being laid for institutional support.

Once again, the lesson learned is that planning pays off. It educates staff about new ideas and services. It prepares library administration for new demands on staff time and funds. It suggests ways to involve the broader community in library endeavors. Investing time in investigation, conversations and synthesize gives us a solid foundation to recommend this new service that will need administrative and financial support.

Acknowledgements:
I thank the other members of the OSU Libraries Task Force on Institutional Repositories for their assistance with this project. They included Mike Boock (Head of Technical Services), Anne Christie (Life Sciences Librarian), Jeremy Frumkin (Gray Chair for Innovative Library Services), Larry Landis (University Archivist) and Laurel Kristick (Electronic Resources Librarian). Also, I thank the OSU Library Administration for supporting my sabbatical so I had the time to focus on this project.

References:

Barton, J., Currier, S. & Hey, J.M.N. 2003. Building quality assurance in metadata creation: An analysis based on the Learning Objects and e-Prints communities of practice. *Dublin Core 2003 Proceedings*. [Online]. Available: http://www.siderean.com/dc2003/201_paper60.pdf [Accessed: 29 July 2004].

Barton, M. R. & Walker, J. H. 2002. *MIT Libraries' DSpace Business Plan Project: Final Report to the Andrew W. Mellon Foundation July 2002*. [Online]. Available: https://libraries.mit.edu/dspace-fed-test/implement/mellon.pdf [Accessed: 29 July 2004].

Corporation for National Research Initiatives. 2003. *The Handle System.* [Online]. Available: http://www.handle.net/ [Accessed: 29 July 2004].

DSpace Federation. 2004. DSpace. [Online]. Available: http://www.dspace.org/ [Accessed: 11 August 2004].

Fox, E. A., Eaton, J.L., McMillan, G., Neill, A. Kipp, N.A., Weiss, L., Arce, E. & Guyer, S. 1996. National Digital Library of Theses and Dissertations: a scalable and sustainable approach to unlock university resources." *D-Lib Magazine* September 1996. [Online]. Available: http://www.dlib.org/dlib/september96/theses/09fox.html#sust [Accessed: 29 July 2004].

Lynch, C. 2003. *Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age.* Association of Research Libraries. 226:1-7. [Online]. Available: http://www.arl.org/newsltr/226/ir.html [Accessed: 29 July 2004].

McGovern, N. Y. 2003. *Preservation in the Age of Repositories.* Lecture at the RLG Member's Forum, December 8-9, 2003. [Online]. Available: http://www.rlg.org/events/haveandhold2003/mcgovern.pdf [Accessed: 29 July 2004].

Nixon, W. J. 2002. The Evolution of an institutional e-prints archive at the University of Glasgow. *Ariadne* 32 [Online]. Available at: http://www.ariadne.ac.uk/issue32/eprint-archives/ [Accessed: 29 July 2004].

Publisher and Library/Learning Solutions (PALS). 2004. *Pathfinder Research on Web-based Repositories: Final Report.* Mark Ware Consulting Ltd., Bristol, U.K. [Online]. Available at: http://www.palsgroup.org.uk/ [Accessed: 29 July 2004].

SPARC & Crow, R. 2002. *Sparc Institutional Repository Checklist and Resource Guide.* SPARC: Washington. D.C. [Online]. Available: http://www.arl.org/sparc/IR/IR_Guide.html [Accessed: 29 July 2004].

Suber, P. 2004. The promise of "Open Access" publishing. *Chronicle of Higher Education Colloquy Live, January 29, 2004.* [Online]. Available: http://chronicle.com/colloquylive/2004/01/openaccess/ [Accessed: 11 August 2004].

Webster, J., Boock, M., Christie, A., Landis, L., Kristick, L & Frumkin, J. 2004. *Implementation of a new library/university service: Oregon State University's Institutional Repository.* Oregon State University Libraries: Corvallis, Oregon.